Towards Morse Theory for Point Cloud Data

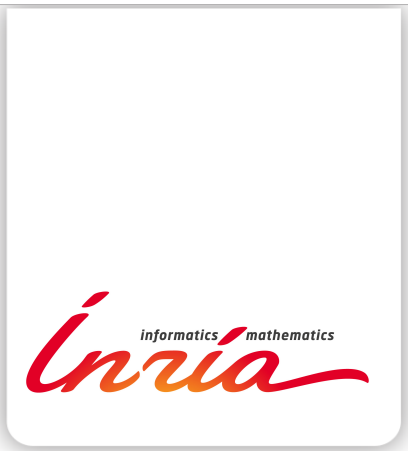
Frédéric Cazals, Christian Mueller, Charles Robert, Andrea Roth

Towards Morse Theory for Point Cloud Data

F. Cazals and C. Mueller and C. Robert and A. Roth

RESEARCH REPORT
N° 8331
July 2013
Project-Team ABS

ISSN 0249-6399 ISRN INRIA/RR--8331--FR+ENG



Towards Morse Theory for Point Cloud Data

F. Cazals * and C. Mueller † and C. Robert ‡ and A. Roth §

Project-Team ABS

Research Report n° 8331 — July 2013 — 34 pages

Abstract:
Morse theory provides a powerful framework to study the topology of a manifold from a function defined on it, but discrete constructions have remained elusive due to the difficulty of translating smooth concepts to the discrete setting.
Consider the problem of approximating the Morse-Smale (MS) complex of a Morse function from a point cloud and an associated nearest neighbor graph (NNG). While following the constructive proof of the Morse homology theorem, we present novel concepts for critical points of any index, and the associated Morse-Smale diagram.
Our framework has three key advantages. First, it requires elementary data structures and operations, and is thus suitable for high-dimensional data processing. Second, it is gradient free, which makes it suitable to investigate functions whose gradient is unknown or expensive to compute. Third, in case of under-sampling and even if the exact (unknown) MS diagram is not found, the output conveys information in terms of ambiguous flow, and the Morse theoretical version of topological persistence, which consists in canceling critical points by flow reversal, applies.
On the experimental side, we present a comprehensive analysis of a large panel of bi-variate and tri-variate Morse functions whose Morse-Smale diagrams are known perfectly, and show that these diagrams are recovered perfectly.

In a broader perspective, we see our framework as a first step to study complex dynamical systems from mere samplings consisting of point clouds.

Key-words: Morse theory, Morse homology, point cloud data, multi-scale analysis, energy landscapes

* Inria Sophia-Antipolis
† ETH Zurich
‡ CNRS
§ Inria Sophia-Antipolis

RESEARCH CENTRE
SOPHIA ANTIPOLIS – MÉDITERRANÉE

2004 route des Lucioles - BP 93
06902 Sophia Antipolis Cedex

Vers une théorie de Morse pour les nuages de points

Résumé : La théorie de Morse fournit un formalisme puissant pour étudier une variété à partie d'une fonction definie sur celle-ci, mais généraliser au cas discret les constructions connues en topologie différentielle est une problématique ouverte.

Considérons le problème consistant à approximer le diagramme de Morse-Smale (MS) d'une fonction à partir d'une graphe de plus proches voisins (NNG) défini sur un nuage de point échantillonant cette fonction. En suivant la construction du theoreme central de l'homologie de Morse, nous présentons des concepts nouveaux de points critiques, ainsi que le diagramme de Morse-Smale associé.

Notre canevas présente trois avantages clefs. Tout d'abord, les primitives utilisées relèvent d'algorithmes et structures de données élémentaires, adaptées au traitement de données en grande dimension. Ensuite, la connaissance du gradient de la fonction étudiée n'est pas requis. Enfin, en cas de sous-echantilonnage et même si le diagramme de MS exact n'est pas retrouvé, des informations sur des *ambiguités* de flot sont mises en évidence. Du point de vue expérimental, nous présentons une analyse exhaustive de fonctions 2D et 3D dont les diagrammes de MS sont connus, et parfaitement retrouvés.

Dans un registre plus général, nous pensons que notre canevas est un premier pas dans la perspective de l'étude de systèmes dynamiques à partir de nuages de points.

Mots-clés : Théorie de Morse, Homologie de Morse, nuages de points, analyse multi-echelle, surfaces d'énergie

Contents

Morse theory provides a powerful framework to study the topology of a manifold from a function defined on it, but discrete constructions have remained elusive due to the difficulty of translating smooth concepts to the discrete setting.

Consider the problem of approximating the Morse-Smale (MS) complex of a Morse function from a point cloud and an associated nearest neighbor graph (NNG). While following the constructive proof of the Morse homology theorem, we present novel concepts for critical points of any index, and the associated Morse-Smale diagram.

Our framework has three key advantages. First, it requires elementary data structures and operations, and is thus suitable for high-dimensional data processing. Second, it is gradient free, which makes it suitable to investigate functions whose gradient is unknown or expensive to compute. Third, in case of under-sampling and even if the exact (unknown) MS diagram is not found, the output conveys information in terms of ambiguous flow, and the Morse theoretical version of topological persistence, which consists in canceling critical points by flow reversal, applies.

On the experimental side, we present a comprehensive analysis of a large panel of bi-variate and tri-variate Morse functions whose Morse-Smale diagrams are known perfectly, and show that these diagrams are recovered perfectly.

In a broader perspective, we see our framework as a first step to study complex dynamical systems from mere samplings consisting of point clouds.

1 Introduction

1.1 Morse Theory and Morse Homology

From differential topology to algorithms. In the introduction of his seminal memoir *Analysis situs*, published in 1895, Henri Poincaré was advocating the investigation of the *relative position of points, lines and surfaces, regardless of their size*, to understand objects beyond the third dimension. While these ideas are recognized to have contributed to the emergence of modern topology, it is probably fair to admit that the study of the variation of a function defined on a topological space, which complements purely topological considerations and is the core of Morse theory, has proved an equally fertile field [Mil63, Bot88]. Morse theory is indeed concerned with the study of the connexion between a function defined on a manifold and the topology of that manifold, and Morse related constructions are key to countless applications, including morphological analysis of 3D shapes in general [BDFF$^+$08] and molecular shapes in particular [CCL03, TN11], image segmentation via the so-called watershed transform [RM00], study of so-called energy landscapes in biophysics [Wal03], high-dimensional data mining [GBPW10], clustering [CadOS11], etc.

On the one hand and from a purely mathematical standpoint, Morse theory [Mil63] and Morse homology [BH04] classically fall in the realm of differential topology and homology theories. In particular, homology calculations can be carried out thanks to the so-called Morse-Smale-Witten chain complex which encodes incidences between critical points whose indices differ by one [BH04]. On the other hand, given the prominence of applied areas which have borrowed upon such ideas, it is especially interesting to see which routes which have been followed to implement Morse related ideas in a computer, given that the data manipulated are inherently discrete at some point. In the following, we briefly list the main approaches, and discuss the key difficulties faced along the way.

Morse theory based on simulation of differentiability. Mimicking the smooth setting, e.g. by linearly interpolating a function defined on the vertices of a simplicial complex has also been

investigated [EHZ01]. The construction culminates with the Morse-Smale complex, namely critical points, their connections, and their (un)stable manifolds. However, the approach is bound to low dimensional data due to the exponential memory footprint of such complexes, and coping with degeneracies such as coalescing critical points remains a challenge.

Morse theory based on arbitrary-precision arithmetic. Given the differential nature of classical Morse theory, it is tempting to design generic algorithm taking for granted arbitrary-precision arithmetic. While such algorithms can be certified under appropriate assumptions, see [CVY12] for the 2D case, they seem of limited practical use due to the stringent assumptions on the numerics.

▷*Discrete Morse theory.* Motivated by the role of so-called CW complexes in Morse theory, Forman developed a purely combinatorial Morse theory, where both the critical elements and the flow operator are defined in terms of cells—practically in a cell or simplicial complex [For98]. However, defining optimal discrete Morse function is in general NP-hard [LLT03], and greedy algorithms, albeit conceptually simple, do not provide intrinsic constructions [CCL03]. Forman's Morse functions were recently used to propose a version of Morse homology [RWS11], which equips the connections of the Morse-Smale complex with the multiplicity information required to form the Morse-Smale-Witten chain complex, from which the homology is computed. Practically, the underlying cell complex being a cubical complex [KMM04, WCV12]. However, for data originating from a smooth function sampled on a 3D grid, relating the critical cells to the critical points of the smooth function is an open question.

Morse theory in the context of Voronoi diagrams. Remarkably, functions used to define affine Voronoi diagrams and α-shapes [Ede03, BDGJ08] have proved amenable to Morse-theoretic constructions, in particular the flow complex [GJ03]. On the positive side, the flow complex enjoys the same properties as the CW complex associated with a classical Morse function, it proved important for reconstructing (non-manifold) shapes [DGRS08, CCS11], and it is also amenable to calculations (critical points and their incidence) in high dimension [GK13]. However, computing the (uns)stable manifolds remains challenging even in 3D [CPP08], letting alone the size of the complex.

Morse theory for point cloud data. It is only recently that Morse theoretical constructions on point cloud data were investigated. For applications to distributed sensor networks analysis, a 2D Morse-Smale decomposition was proposed in [ZSG09]. The algorithm handles domains of arbitrary topology, but the construction used to identify critical points (cut locus) is bound to 2D, and no comparison is reported against certified MS diagrams of smooth functions. The topography of graphs embedded in dD is studied in [WLY12]. But the notion of critical event proposed in not discussed in the context of smooth functions, and is actually suffers from a *contamination effect* (see below, description of step 1).

Contributions. This work makes a stride towards an algorithmic version of Morse theory enjoying the following key properties:

- Ease of implementation and ability to handle high-dimensional data.
- Amenability to simple interpretations for data associated with smooth Morse functions.
- Encapsulation of Morse homology calculations.

Guided by the first wish, our approach is based on the analysis of point cloud data, that is, we only assume as input a nearest neighbor graph connecting the samples of a point cloud sampling a smooth Morse function, and we only resort to sorting and traversals on graphs. To match the second wish, a *critical point* is plainly a sample approximating the exact critical point

of the function studied. Similarly, the (un)stable manifold $W^s(p)$ ($W^u(p)$) of a critical point p will consist of a sub-graph of the input graph.

Central to our constructions is the notion of *bifurcation*, which aims at capturing ambiguities of the flow, such as those found on the compactification of a region defined as the intersection $W^u(p) \cap W^s(q)$ of unstable and stable manifolds for a smooth Morse function. While these constructions are not backed-up by theorems at this stage, we believe that our work makes two important contributions. First, we provide insights on subtle behaviors of discrete flow lines, which are not faced in the smooth setting. Second, we report correct Morse-Smale diagrams for challenging Morse functions in 2D and 3D, the systematic comparison of a smooth Morse-Smale diagram against its approximation being undertaken for the first time in this work, to the best of our knowledge.

1.2 Notations

Graphs and related constructions. A set denoted by a plain capital letter; a set of sets is denoted with a calligraphic font ie $\mathcal{F} = \{F_i\}$.

A graph defined by its vertex and edge sets is denoted in sans serif font, e.g. $\mathsf{G} = (V[\mathsf{G}], E[\mathsf{G}])$. An *induced sub-graph* G' of G is a sub-graph of G such that for all pairs of vertices (u, v) of G', (u, v) is an edge of G' iff it is an edge of G. The sub-graph induced by a vertex set $T = V[\mathsf{G}'] \subset V[\mathsf{G}]$ is defined similarly: (u, v) is an edge of G' iff it is an edge of G. A Hasse diagram is a directed acyclic graph. In a directed graph, the nodes pointed by outgoing (resp. incoming) edges are called the successors (resp. ancestors) of that node.

We consider a set of samples denoted $P = \{p_i\}_{i=1,\ldots,n}$, each endowed with a real function value denoted by $f(p_i)$. We assume that these samples are connected by a symmetric nearest neighbors graph (NNG), denoted G. The *lower link of a sample* consists of its neighbors in the NNG having a lower function value.

Discrete flow operator. We use the NNG connecting the samples to define a *descending pseudo-gradient graph* G^- by linking each sample p_i to a sample q_i, thus approximating the negative gradient of the function f. (Several options are explored in section 3.2.) A sample p_i is called *regular* in G^- if it has an outgoing edge, and *singular* otherwise. Singular vertices are *local minima*. The discrete *flow operator* is the operator iteratively following the neighbor (below in G^-, above in G^+) of a sample. Using this flow operator, the origin and destination of a sample p_i are denoted $\alpha^-(p_i), \omega^-(p_i)$ if we follow the vectors of G^-. Note in particular that a local minimum is characterized by $\omega^-_{|\mathsf{G}}(p_i) = p_i$.

The previous notions generalize for any induced sub-graph of G. Given a set of samples $H \subset V[\mathsf{G}]$, the *induced descending pseudo-gradient graph* $\mathsf{G}^-_{|H}$ is the pseudo-gradient graph defined on the sub-graph of G induced by the set $V[\mathsf{H}]$. Note that this graph is different from the sub-graph of G^- induced by the set $V[\mathsf{H}]$. (If a node from H flows to a node of $V[\mathsf{G}]\backslash H$, it is orphan in the latter graph.) Likewise, the destination of a sample under the descending flow is denoted $\omega^-_{|\mathsf{H}}(p_i)$, etc.

Morse theory. An index k critical point is denoted $\sigma^{(k)}$, and its (un)stable manifolds are respectively $W^s(\sigma^{(k)})$ and $W^u(\sigma^{(k)})$. The index of a critical point p is denoted λ_p. The sub-level set of a function f up to level h is denoted $M[f]_h$ or M_h for short if the function is clear.

Conventions for artwork. For screen-shots, the colors associated with critical points are as follows: index 0 i.e. local minimum: purple; index 1 saddle: blue; index 2 saddle: orange; index 3 i.e. local maximum: red.

2 Construction: Overview

2.1 Ingredients of Morse homology

Our framework to study a sampled Morse function is guided by the smooth setting, and by the so-called Morse homology theorem, which allows computing the homology of the underlying manifold from the incidences between pairs of critical points whose indices differ by one. The key ingredients of Morse homology are [BH04]:

- The flow lines (i.e.integral curves) associated with the negative gradient of the Morse function $-\nabla f$, and the associated orbits, that is the space of flow lines quotiented out by additive reparameterizations.
- The stable and unstable manifolds of the critical points, from which one defines regions of homogeneous flow $W(q,p) = W^u(p) \cap W^s(q)$. When $W(q,p) \neq \emptyset$, this set is used to define a partial order $q \succ p$ between pairs of critical points connected by an orbit. The critical points of f and this partial order define the *phase diagram* of the Morse function.
- The compactification $\overline{W(q,p)}$ of of $W(q,p)$. It can be shown that the *boundary* of $\overline{W(q,p)}$ consists of flow lines going through critical points intermediate between q and p. These flow lines are called *boundary* flow lines.
- The *moduli space* $\mathcal{M}_{q,p} = W(q,p)/\mathbb{R}$, namely orbits joining q to p. In particular, when $\lambda_q = k$ and $\lambda_p = k-1$, $\mathcal{M}_{q,p}$ has dimension zero, i.e. consists of a finite number of orbits.
- The so-called Morse-Smale-Witten chain complex (MSW) connecting pairs of critical points q and p with $\lambda_q - \lambda_p = 1$, together with the signed multiplicities of these orbits, so as to define the boundary operator from which the homology groups are computed. Note that when the multiplicity is omitted, we plainly talk about the Morse-Smale (MS) complex.

2.2 Overview

We aim at capturing the orbits of the moduli spaces $\overline{W(q,p)}$ with $\lambda_q - \lambda_p = 1$. As recalled above, connections between such critical points are found on the boundary of regions of homogeneous flow $W(q,p)$. Because non empty regions $W(q,p) \cap W(q,p')$ precisely delimit two regions of homogeneous flow, our strategy precisely consists of identifying such regions, thanks to *bifurcations*. A sample p_i *bifurcates* if it faces a flow ambiguity in the following sense: selected neighbors of his lower link flow to different index $k-1$ critical points. (Prosaically, two neighboring samples sitting across a ridge on a mountain define a bifurcation provided that they flow to different local minima.)

Albeit intuitive, the notion of bifurcation is difficult to capture: on the one hand and as just discussed, bifurcations are intimately related to flow ambiguities; on the other hand, while orbits are disjoint in the smooth setting, merges and splits are observed between their discrete analogs, due to the discrete nature of the sampling. In particular, the notions of *confluence samples, forks, sticks* and *linkers* to be introduced precisely aim at retaining the proper connections between critical points, which is non trivial due to these merges and splits.

2.3 Key steps

We now present the key steps and constructions, illustrated on Figs. 1 and 2, the details being provided in the supplemental section 6. Due to the aforementioned structure of the MSW chain complex, our construction iteratively builds slices of the MS complex, a slice precisely consisting

of the index k critical points, their unstable manifolds, and the connections to index $k-1$ critical points. To this end, the kth iteration requires (i) the index $k-1$ critical points, (ii) a graph called the global thickening $\mathsf{T}[k-1]$, and (iii) the associated pseudo-gradient graph. For the first iteration ($k=1$), this information consists of the local minima, and the global thickening $\mathsf{T}[0]$ is the NNG, denoted G.

To ease the presentation, for each step, we first present the main ideas, and proceed with subtle issues (identified by the pictogram), whose understanding requires material presented in the supplemental section 6.

▷ STEP 1: **Identifying bifurcating and confluence samples** This step aims at identifying one set $B_i^{(k)}$ of *bifurcating samples*, and one set $C_i^{(k)}$ of *confluence samples*, by processing the vertices of the global thickening $\mathsf{T}[k-1]$ by increasing function value (Fig. 1(A,B)):
– A bifurcating sample is such that two points from its lower link (bifurcation lower link, actually), flow to different index $k-1$ critical points. As discussed above, such a sample is located in a region that lies on the boundary of two regions of homogeneous flow. Critical points will be found amongst bifurcating samples. Through a bifurcating sample p_i, there is a boundary flow line flowing to a $\sigma^{(k-1)}$ by passing through a $\sigma^{(k)}$ first. As an example on Fig. 1, p_i flows to a local minimum by visiting an index one saddle first.
– A confluence sample is a sample such that its entire lower link consists of bifurcating samples. Intuitively, such samples witness the confluence of flow lines, an artifact of the discretization.

These intuitive definitions actually hide one complication: when checking whether a sample p_i is bifurcating, one should use the subset of its lower link consisting of non-bifurcating samples. If this precaution is not taken, because the samples are processed by increasing function value, a contamination effect is observed. That is, the number of bifurcating samples increases as a function of the distance to local minima of $\mathsf{T}[k-1]$ (the distance is counted in numbers of edges in the global thickening). This accounts for the notion of *bifurcating lower link*, which is a subset of the lower link for $k \geq 2$ (supplemental section 6.3). We also define the global stream $\mathsf{B}[k]$, namely the sub-graph of $\mathsf{T}[k-1]$ induced by the bifurcating samples. Note that the remaining samples, which are regular, are of no interest for further processing, since they belong to the stable manifolds of the index $k-1$ critical points.

▷ STEP 2: **Computing anchors and local thickenings** We wish to partition the samples of $\mathsf{B}[k]$ into equivalence classes with respect to tuples of index $k-1$ critical points (Fig. 1(C)). Consider such a subset, which we call a k-anchor, denoted $A^{(k)}$. All the samples from $\mathsf{B}[k]$ having the same anchor define the *local thickening* $\mathsf{T}(A^{(k)})$ of this anchor. Intuitively, a local thickening is found across two (ore more) regions of homogeneous flow. Intuitively, if a local thickening has a unique local minimum, it represents the stable manifold of the critical point identified with this local minimum. If not, it represents the union of the stable manifolds associated to critical points sharing the same successors along the flow, identified by the critical points $\sigma^{(k)}k-1$ defining its anchor.

▷ STEP 3: **Computing monotonic sections and their incidence graph** A local thickenings witnesses bifurcations with respect to the same index $k-1$ critical points, regardless of the variation of the function itself. We therefore compute a covering of $\mathsf{T}(A^{(k)})$ with a finite set of monotonic sections $\mathsf{M}^{(\alpha,\omega)}(\mathsf{T}(\mathsf{A}^{(\mathsf{k})}))$, each of them consisting of samples sharing the same α and ω limits within $\mathsf{T}(A^{(k)})$ (Fig. 1(C)). Furthermore, the edges belonging to the global stream but not to any monotonic section are used to build the graph $G_M^{(k)}$ whose nodes are monotonic sections (but so-called linkers, see below), with one edge for two incident such sections.

Confluence samples do not bifurcate by construction (Figs. 6 and 7, but monotonic sections consisting of such samples (except their boundary) are important since following them downstream brings to regions where monotonic sections associated to different anchors split. Such monotonic sections are called *linkers*, and are excluded from $G_M^{(k)}$ (supplemental section 6.5). In fact, linkers allow maintaining the connections between critical points of index k and index $k-1$ (supplemental section 6.6). Note that linkers do not contribute critical points. Note also that linkers account for the definition of bifurcating sets using confluence samples, as seen in step 1.

▷ STEP 4: **Simplifying the incidence graph of monotonic sections** We wish to transform the incidence graph $G_M^{(k)}$ into a simplified graph $\overline{G}_M^{(k)}$ compatible with the MSW chain complex, by retaining edges $(\mathsf{M}_\mathsf{i}^{(k)}, \mathsf{M}_\mathsf{j}^{(k)})$ between monotonic sections which potentially contribute critical points whose indices differ by one, and such that the value of the function on $\mathsf{M}_\mathsf{i}^{(k)}$ less than the value of the function on $\mathsf{M}_\mathsf{j}^{(k)}$ (recall that the function value decreases along a flow line). Because each monotonic section has a defining anchor $A^{(k)}$ (a tuple of index $k-1$ critical points), and since anchors enjoy a partial order for the inclusion, the edition of $G_M^{(k)}$ into $\overline{G}_M^{(k)}$ is based on two partial orders: the inclusion partial order $\prec_i$ just discussed, and a function value related partial order denoted $\prec_h$, which qualifies the relative height of monotonic sections which are connected by a path in $G_M^{(k)}$. Given two elements to be compared, two partial orders yield five different cases: the two elements can be compared (two cases), only one partial order applies (two cases), the elements cannot be compared by any of the two partial orders (one case). Accordingly, the edges of $G_M^{(k)}$ are classified as *canonical, composite, H-non-I, I-non-H*, or *non-hierarchical*. The graph $G_M^{(k)}$ is then edited into $\overline{G}_M^{(k)}$ by keeping canonical edges only, so that $\overline{G}_M^{(k)}$ precisely consists of the edges (and associated incident nodes i.e. monotonic sections) satisfying the following two conditions, which define the partial order denoted $\prec_{ih}$:

$$\mathsf{M}_\mathsf{i}(A_u^{(k)}) \prec_i \mathsf{M}_\mathsf{j}(A_v^{(k)}) \text{ and } \mathsf{M}_\mathsf{i}(A_u^{(k)}) \prec_h \mathsf{M}_\mathsf{j}(A_v^{(k)}). \tag{1}$$

The difficulty consists of handling the edges which are not canonical. These edges are indeed related to artifacts of the discretization in general. For example, composite edges are related to the existence of monotonic sections called *forks* and *sticks* (supplemental Fig. 6), namely incident monotonic sections such that the two partial orders $\prec_h$ and $\prec_i$ *disagree*. When detected at stage k, these sections actually yield confluence samples at the $k+1$th iteration. Other situations account for the three remaining types of edges (supplemental section 6.6).

▷ STEP 5: **Computing the kth slice of the Morse-Smale complex** The graph $\overline{G}_M^{(k)}$ is used to build the kth slice $L_{MS}^{(k)}$ of the Morse-Smale complex. First, index k critical points $\sigma^{(k)}$ are identified: each monotonic section of the graph $\overline{G}_M^{(k)}$ which is minimal with respect to the partial order $\prec_{ih}$ contributes its local minimum as an index k critical point. Second, the connections between each index k critical point $\sigma^{(k)}$ and its successors are set. Consider the anchor associated with the monotonic section defining $\sigma^{(k)}$: we create one connexion between $\sigma^{(k)}$ and each index $k-1$ critical point of this anchor. Third, stable manifolds of index $\sigma^{(k-1)}$ are computed. More precisely, $W^s(\sigma^{(k)})$ consists of the samples of $\mathsf{B}[k]$ together with the samples from linkers discovered during the current iteration, which flow to $\sigma^{(k)}$.

▷ STEP 6: **Computing the global thickening for the next iteration** The final step consists in preparing the global thickening $\mathsf{T}[k]$ to be used during the next iteration. The vertex set of this graph consists of the vertices of the monotonic sections in $\overline{G}_M^{(k)}$.

 The graph $\mathsf{T}[k]$ is not an induced sub-graph of $\mathsf{T}[k-1]$, due to the removal of edges in $\mathsf{T}[k-1]$ responsible for the non-hierarchical incidences in $G_M^{(k)}$.

3 Experiments

3.1 Benchmark

Rationale. To test our construction, we selected challenging functions of two and three variables whose Morse-Smale diagrams are known exactly. (Equations, critical points, and illustrations are presented in the supplemental section 7.) We devoted a particular attention to polynomial functions, whose critical points can be certified using *real solving* tools. More precisely, for a multi-variate polynomial P in d variables, one can form the system equating all partial derivatives to zero, and solve it if it is zero-dimensional—the solution set consists of isolated points. This task is performed by the Maple function `RootFinding[Isolate]`, which combines four ingredients, namely (i) a Gröbner basis calculation with the F4 algorithm [Fau99], (ii) a Rational Univariate Representation (RUR) calculation [Rou99], (iii) a certified isolation of the roots of the polynomial produced by the RUR [RZ03], and (iv) multi-precision interval arithmetic to report boxes isolating the solutions of the system [RR05]. Upon isolating the roots, one evaluates the Hessian matrix at each critical point, and gets the index from the number of negative eigenvalues.

In discussion these examples, the signature of a function is the number of critical points of each index.

Functions of two variables. The 2D *Himmelblau* function is a quartic polynomial defining a simple multimodal landscape, yet challenging for optimization algorithms since the four global minima are at the same height. Its signature is $(4, 4, 1)$, and all critical points can be certified using the aforementioned real solving tools. The *2D Rastrigin* function [MSB91] is a challenging case due to the cosine modulation of a parabolic function. In the domain $[-5.12, 5.12]^2$, its signature is $(121, 220, 100)$. We also created two terrains (Gauss6A, Gauss6B), each defined as a sum of six Gaussians. In particular, the former illustrates the case where two index one saddles are connected to the same two local minima.

Functions of three variables. Since we did not find any function with certified sets of critical points the literature, we used *quartic polynomial* functions whose critical points were certified using real solving. The associated MS complex was established by visual inspection, by tracking connections between sub-level sets.

To probe our construction beyond smooth functions, we used the Euclidean distance to a 3D point cloud, resulting in a Morse-Smale diagram known as the flow complex [GJ03]. For that case too, certified MS diagram can be obtained using multi-precision rational numbers [CPP08].

3.2 Protocol

Implementation. We implemented our algorithm, whose pseudo-code is provided in the supplemental section 6.1, in generic C++. The NNG connecting the samples being taken for granted, each iteration requires sorting steps and graph traversal algorithms. That the worst-case complexities are linear in the size of the graphs manipulated (global thickening, local thickenings, monotonic sections and their incidence graph).

The algorithm works in any dimension, and can be instantiated with a variety of points and distances (points in a Euclidean space under the Euclidean distance, points representing molecular conformations under the least root mean square deviation, etc).

Nearest Neighbor Graph and pseudo-gradient graph. To focus on intrinsic properties of the construction and avoid difficulties related to an uneven sampling, we used point clouds sampled on a grid superimposed to the 2D or 3D domain of interest. Thanks to this grid sampling, the neighborhood of a sample was defined as its set of l neighbors, with $l = \sum_{k=0}^{d-1} 2^{d-k} \binom{d}{k}$.

To defined the PSG, we plainly connected each sample p_i to its neighbor q_j defining the steepest slope, amidst the l candidates.

3.3 Results

Timings. All the datasets processed, up to one million of points, required less than five minutes on a desktop computer (one core Intel Xeon(R) CPU X5482 at 3.2GHz, running Linux Fedora core 17).

Smooth functions of two variables. Exact results (critical points and their connections) are obtained for all the functions under consideration, for samplings as small as 100^2. See Fig. 3.

Smooth functions of three variables. We obtained correct results for samplings in the range 25^3 to 50^3. One particular difficulty we encountered is linked to the presence of *open local thickenings* (supplemental section 6.4), which rarely produced redundant critical points, filtered out as explained below. See Fig. 4.

Non smooth functions: the 3D flow complex. Correct results were also obtained, despite the non-smoothness of the distance function. At times, several critical points were also observed within monotonic sections, and were filtered out as explained below. A simple case is illustrated on Fig. 2.

Filtering out spurious critical points. By construction, each monotonic section is defined with respect to one local minimum and maximum of a local thickening (supplemental section 6.7). In rare cases, redundant local minima (and also maxima) located near one another were observed within a local thickening, and all but one were filtered out. To see how, consider two local minima p and q having a common ancestor r in the local thickening, with p nearest to r in function value (w.l.o.g.). In all the cases observed, the distance between p and r was at most three edges. To filter out p, we plainly revert the flow from p to r, which is the usual strategy to run persistence from the Morse-Smale diagram [CCS11].

4 Discussion and Outlook

This paper presents a framework to compute an approximation of the Morse-Smale diagram of a function, from a point cloud sampling the domain over which the function is defined. Central is the identification of *bifurcating* samples, i.e. samples facing ambiguities in terms of discrete flow defined on a nearest neighbor graph connecting the input points. From an algorithmic standpoint, our framework has three key advantages. First, it requires elementary data structures and operations, and is thus suitable for high-dimensional data processing. Second, in the context of non convex optimization, it is *gradient free*, which makes it suitable to investigate functions whose gradient is unknown or expensive to compute. Third, in case of under-sampling and even if the exact (unknown) MS diagram is not found, the output conveys information in terms of ambiguous flow, and the Morse theoretical version of topological persistence, which consists in canceling critical points by flow reversal, applies. From an applied standpoint, we present exact MS diagram for various challenging functions of two and three variables (including smooth functions and distance functions), a miles-stone first achieved in this paper to the best of our knowledge.

Despite these achievements, we foresee both theoretical and applied extensions.

On the theoretical side, two directions seem particularly promising. The first one is related to the investigation of sufficient conditions on the sampling, so as to ensure an isomorphism between the MS diagram computed and the exact one. Under such hypothesis and upon decorating the edges of the MS diagram with their (signed) multiplicity, our construction would allow computing the (persistent) homology of the underlying manifold using the Morse-Smale-Witten boundary operator — a very attractive perspective since this boundary operator handles a CW complex with one cell per critical point. The second one deals with the extension of our framework to dynamical systems beyond gradient vector fields, so as to detect invariant sets such as attractors consisting of fixed orbits. We believe in particular that some of the notions introduced in our work, related to the confluence of flow lines, will prove useful in this respect.

On the applied side, given the ubiquity of dynamical systems in science and engineering, our constructions should find various applications, and we mention four of them. First, we plan to use it in the context of Hamiltonian systems, where phase transitions are typically investigated through the variation of the Euler characteristic χ of the phase space. Our approximate MS complex, extended to the MSW chain complex as just discussed, clearly contain more information. Second, our framework actually handles points sampling a space of arbitrary topology — not just a contractible domain. We thus plan to use it to investigate sampled energy landscapes with non trivial topologies, as frequently encountered for molecular system that do not visit their entire configuration space (intuitively, orphan regions correspond to locii where the potential energy is prohibitively high). Third, we plan to work on the design of collective coordinates of macro-molecular systems, namely coordinates accounting for low frequency - large amplitude of molecular motions. The design of such coordinates is typically based on the identification of (significant) stable minima of the energy landscape of the system, and our ability to report all transitions paths connecting two local minima across index one saddles is appealing to handle cases featuring saddles of comparable heights. Fourth, given the excellent performances observed on tri-variate functions sampled on a grid, we plan to use our tool in the context of 3D image analysis, and also to analyze general functions — the latter in the context of scientific software such as Maple or Mathematica.

Acknowledgments. T. Dreyfus is acknowledged for his help with the software development.

This work has partially been supported by the Computational Geometric Learning STREP project of the EC 7th Framework Programme (EC contract No. 255827).

References

[BDFF+08] S. Biasotti, L. De Floriani, B. Falcidieno, P. Frosini, D. Giorgi, C. Landi, L. Papaleo, and M. Spagnuolo. Describing shapes by geometrical-topological properties of real functions. *ACM Computing Surveys (CSUR)*, 40(4):12, 2008.

[BDGJ08] K. Buchin, T. Dey, J. Giesen, and M. John. Recursive geometry of the flow complex and topology of the flow complex filtration. *Computational Geometry - Theory and Application*, 40(2):115–137, 2008.

[BH04] A. Banyaga and D. Hurtubise. *Lectures on Morse Homology*. Kluwer, 2004.

[Bot88] R. Bott. Morse theory indomitable. *Publications Mathématiques de l'IHÉS*, 68(1):99–114, 1988.

[CadOS11] F. Chazal, L.J. Guibas an dS.Y. Oudot, and P. Skraba. Persistence-based clustering in riemannian manifolds. In *ACM SoCG*, pages 97–106, 2011.

[CCL03] F. Cazals, F. Chazal, and T. Lewiner. Molecular shape analysis based upon the Morse-Smale complex and the Connolly function. In *ACM Symposium on Computational Geometry*, San Diego, 2003.

[CCS11] F. Cazals and D. Cohen-Steiner. Reconstructing 3d compact sets. *Computational Geometry Theory and Applications*, 45(1-2):1–13, 2011.

[CPP08] F. Cazals, A. Parameswaran, and S. Pion. Robust construction of the three-dimensional flow complex. In *ACM SoCG*, pages 182–191, 2008.

[CVY12] A. Chattopadhyay, G. Vegter, and C. Yap. Certified computation of planar morse-smale complexes. In *ACM SoCG*, 2012.

[DGRS08] T. Dey, J. Giesen, E. Ramos, and B. Sadri. Critical points of the distance to an epsilon-sampling of a surface and flow based surface reconstruction. *Int. J. of Computational Geometry and Applications*, 18(2):29–62, 2008.

[Ede03] H. Edelsbrunner. Surface reconstruction by wrapping finite point sets in space. In B. Aronov, S. Basu, J. Pach, and M. Sharir, editors, *Discrete and Computational Geometry (Ricky Pollack and Eli Goodman Festschrift)*, pages 379–404. Springer-Verlag, 2003.

[EHZ01] H. Edelsbrunner, J. Harer, and A. Zomorodian. Hierarchical Morse complexes for piecewise linear 2-manifolds. In *ACM SoCG*, 2001.

[Fau99] J.-C. Faugère. A new efficient algorithm for computing gröbner bases (f_4). *Journal of Pure and Applied Algebra*, 139(1-3):61–88, 1999.

[For98] R. Forman. Morse theory for cell complexes. *Advances in Mathematics*, 134:90–145, 1998.

[GBPW10] Samuel Gerber, P Bremer, Valerio Pascucci, and Ross Whitaker. Visual exploration of high dimensional scalar functions. *Visualization and Computer Graphics, IEEE Transactions on*, 16(6):1271–1280, 2010.

[GJ03] J. Giesen and M. John. The flow complex: A data structure for geometric modeling. In *ACM SODA*, 2003.

[GK13] J. Giesen and L. Kuhne. A parallel algorithm for computing the flow complex. In *ACM SoCG*, Rio de Janeiro, 2013.

[KMM04] T. Kaczynski, K. Mischaikow, and M. Mrozek. *Computational Homology*. Springer, 2004.

[LLT03] T. Lewiner, H. Lopes, and G. Tavares. Optimal discrete Morse functions for 2-manifolds. *Computational Geometry: Theory and Applications*, 26(3):221–233, 2003.

[Mil63] John W. Milnor. *Morse Theory*. Princeton University Press, Princeton, NJ, 1963.

[MSB91] H. Mühlenbein, M. Schomisch, and J. Born. The parallel genetic algorithm as function optimizer. *Parallel computing*, 17(6-7):619–632, 1991.

[RM00] J. Roerdink and A. Meijster. The watershed transform: Definitions, algorithms and parallelization strategies. *Mathematical morphology*, page 187, 2000.

[Rou99] F. Rouillier. Solving zero-dimensional systems through the rational univariate representation. *Journal of Applicable Algebra in Engineering, Communication and Computing*, 9(5):433–461, 1999.

[RR05] N. Revol and F. Rouillier. Motivations for an arbitrary precision interval arithmetic and the mpfi library. *Reliable Computing*, 11:1–16, 2005.

[RWS11] V. Robins, P. Wood, and A. Sheppard. Theory and algorithms for constructing discrete morse complexes from grayscale digital images. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, (99):1–1, 2011.

[RZ03] F. Rouillier and P. Zimmermann. Efficient isolation of polynomial real roots. *Journal of Computational and Applied Mathematics*, 162(1):33–50, 2003.

[TN11] D. Mathew Thomas and V. Natarajan. Symmetry in scalar field topology. *IEEE Transactions on Visualization and Computer Graphics*, 17(12):2035–2044, 2011.

[Wal03] D.J. Wales. *Energy Landscapes*. Cambridge University Press, 2003.

[WCV12] H. Wagner, C. Chen, and E. Vuçini. Efficient computation of persistent homology for cubical data. *Topological Methods in Data Analysis and Visualization II*, pages 91–106, 2012.

[WLY12] E. Weinan, J. Lu, and Y. Yao. The landscape of complex networks. Technical Report http://arxiv.org/abs/1204.6376, arxiv, 2012.

[ZSG09] Xianjin Zhu, Rik Sarkar, and Jie Gao. Topological data processing for distributed sensor networks with morse-smale decomposition. In *INFOCOM 2009, IEEE*, pages 2911–2915. IEEE, 2009.

5 Artwork

Figure 1 Regions of homogeneous flow, bifurcations, and monotonic sections, illustrated for $d = 2$ and $k = 1$ (A) A region of homogeneous flow $W(\sigma_0^{(2)}, \sigma_1^{(0)}) = W^u(\sigma_0^{(2)}) \cap W^s(\sigma_1^{(0)})$ on a two dimensional manifold. The compactification of this region also involves flow lines running through the intermediate critical points $\sigma_1^{(1)}$ and $\sigma_2^{(1)}$. The circled blue region marks a neighborhood of one of the boundary flow lines, analyzed in B. The orange area marks a region where a confluence of two flow curves respectively descending from $\sigma_2^{(2)}$ and $\sigma_3^{(2)}$ occurs: in the smooth setting, these curves would only meet at the local minimum; in the discrete case, they coalesce before. **(B)** Points in the neighborhood of the boundary flow line of (A) may flow to $\sigma_1^{(0)}$ or to $\sigma_1^{(0)}$. This ambiguity is used to define the the bifurcating samples (red samples), which define the set $B_n^{(0)}$). Each such sample is witnessed by its bifurcation lower link $L_{|\mathsf{G}}^{\rightsquigarrow}(p_i)$ depicted in blue. Note that p_l does not belong to the bifurcation lower link of p_i since it bifurcates itself. **(C)** Samples bifurcating with respect to the same tuples of index $k-1$ critical points, called an anchor $A^{(k)}$, are grouped into a local thickening $\mathsf{T}(A^{(k)})$, and each local thickening is itself partitioned into monotonic regions $\mathsf{M}^{(\alpha,\omega)}(\mathsf{T}(\mathsf{A}^{(\mathsf{k})}))$. Displayed here are the four monotonic regions associated with the anchor $(\sigma_1^{(0)}, \sigma_2^{(0)})$ involving the two local minima. Two monotonic regions associated to the same anchors (resp. two different anchors) are termed interior (resp. exterior) incident.

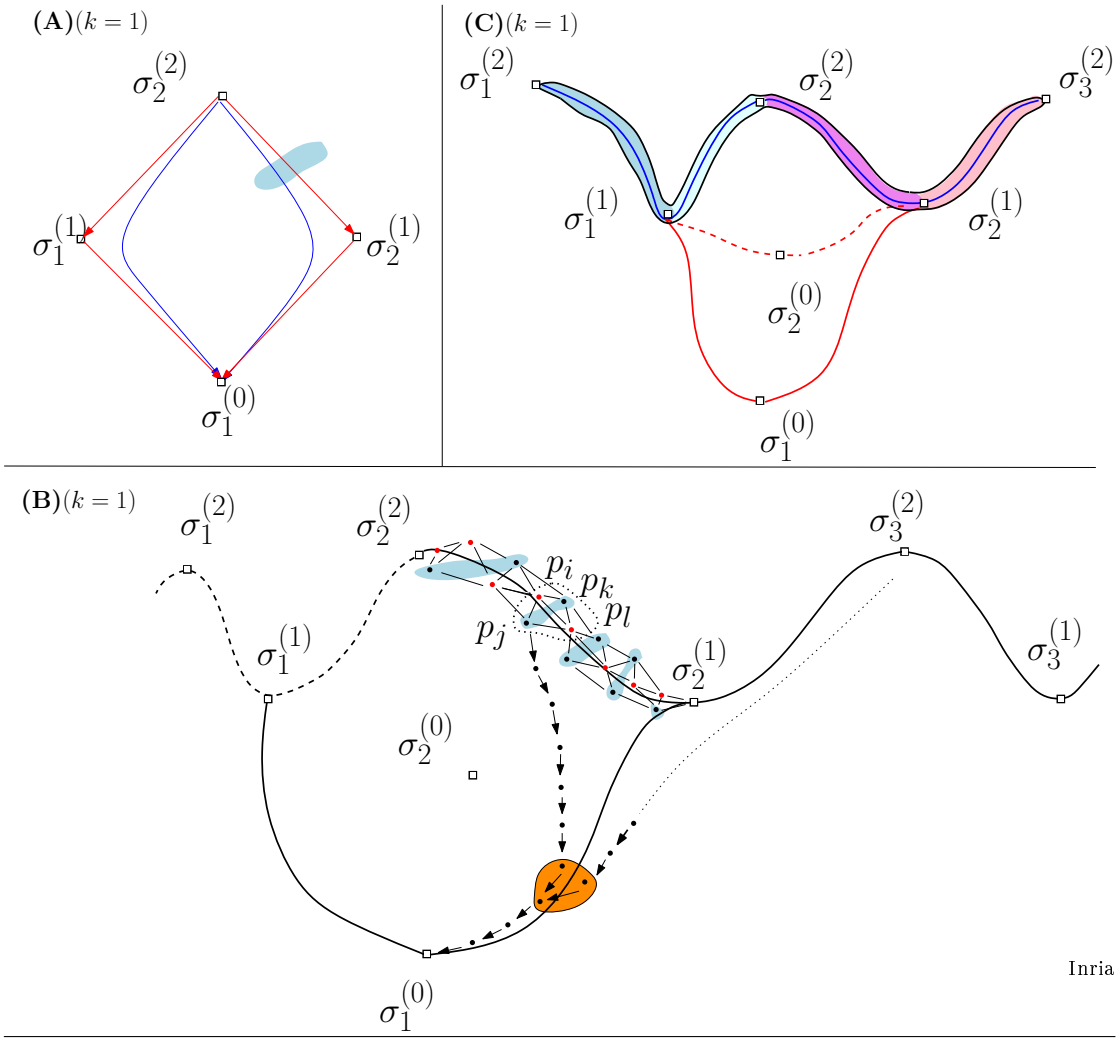


Figure 2 Concepts illustrated for the Euclidean distance function to the vertices of a tetrahedron (Main) Seven critical points $\sigma_i^{(k)}$ are represented: index 0 in purple, index 1 in blue, index 2 in orange, index 3 in red. The three monotonic sections $\mathsf{M}_1^{(1)}, \mathsf{M}_2^{(1)}, \mathsf{M}_3^{(1)}$ partitioning the local thickening of the pair of local minima $(\sigma^{()}, \sigma^{()})$ are also tagged. **(Inset)** Overview of the 7 critical points of the main figure.

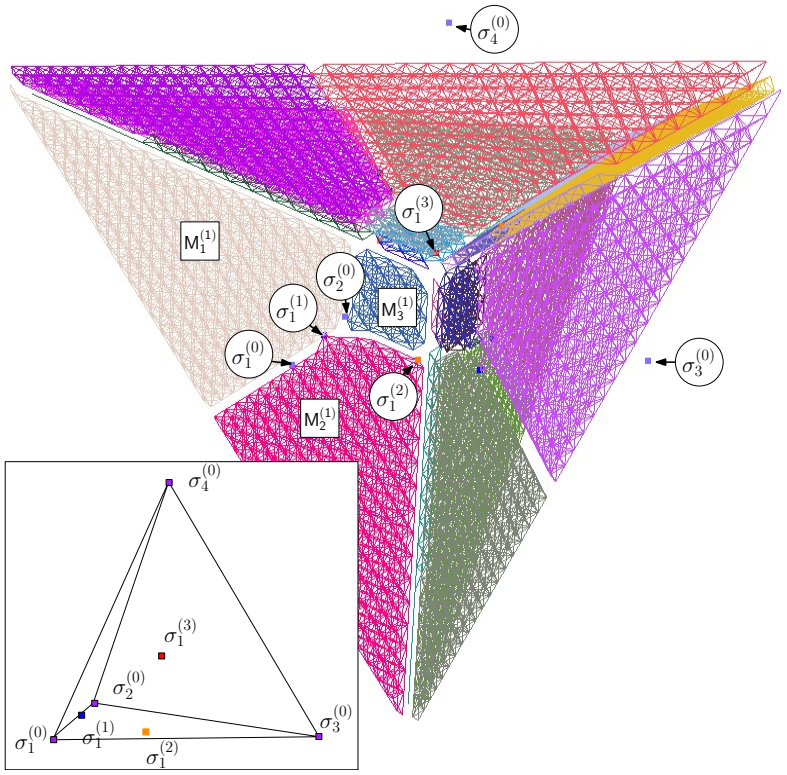
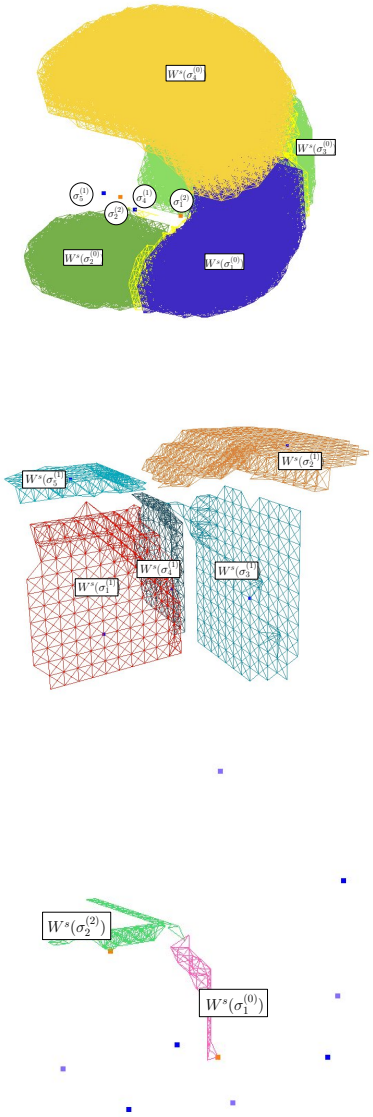


Figure 3 The critical points and manifolds of 2D functions (Top) The Himmelblau function has signature (4,4,1). All minima are at the same critical value. **(Middle)** The Rastrigin function has (121,220,100) as signature. **(Bottom)** The Gauss6A function has (3,5,3). One pair of minima shares three $\sigma^{(1)}$. In the inset we can notice the two $\sigma^{(1)}$ (blue points) that are linked to the same anchor.

Figure 4 Stable manifolds of the critical points of poly1 (Top) Stable manifolds of the four local minima up to the critical value of $\sigma_4^{(1)}$. Note that the $W^s(\sigma_i^{(0)})$ are separated by the global thickening $\mathsf{T}[1]$. **(Middle)** Sub-level set before the apparition of an index 2 critical point. The stable manifolds of $\sigma^{(1)}$ (pink points) and their boundary, the 2-global thickening(yellow) are depicted. **(Bottom)** The stable manifolds of $\sigma^{(2)}$ (orange points) are depicted. However, as the terrain does not have any maximum, these don't intersect.



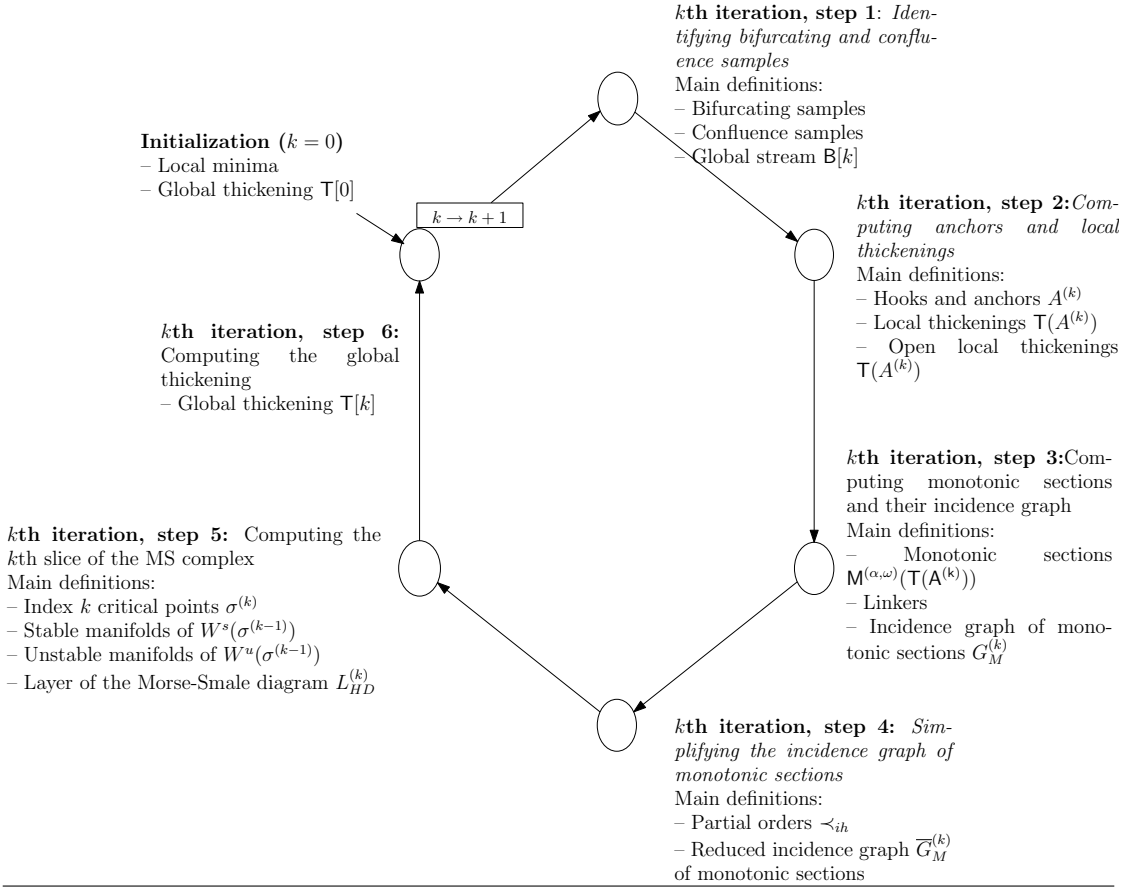
6 Appendix: Details of the Construction

6.1 Algorithm Pseudo-code

The six steps of the construction are recalled on Fig. 5, and the corresponding pseudo-code is presented in Algorithm 1. Complexity-wise, the following comments are in order for the execution of each iteration:

- The samples of $\mathsf{T}[k]$ by increasing function value.
- Computing the alpha and omega limits of each sample within $\mathsf{T}[k]$ is linear in the size of the graph, in the worst-case.
- Computing the anchor of each point requires the omega limit for all samples in its bifurcation lower link.
- Computing local thickenings requires sorting samples using the lexicographic ordering on the tuples defining their anchors.
- Identifying the local minima and maxima in the local thickenings has again linear complexity in the size of these local thickenings, in the worst-case.
- Since incidences between monotonic sections are recorded while creating the monotonic sections, creating the incidence graph $G_M^{(k)}$ is linear in its size.
- All subsequent operations are also linear in the size of $G_M^{(k)}$.

Figure 5 The six steps of our construction, together with the main definitions.



Algorithm 1 Algorithm

(0) Pre-processing
  Compute nearest neighbor graph
  Compute pseudo-gradient graph
  Identify local minima

Build the kth slice of the Morse-Smale complex:

(1) Compute the bifurcating and the confluence samples w.r.t. $\{\sigma^{(k-1)}\}$ in the global thickening
(2) Group samples into equivalence classes, called *local thickenings*
(3a) Decompose each local thickening into monotonic sections
(3b) Construct graph of incidence between all monotonic sections, excepting linkers
(4) Simplify the incidence graph between monotonic sections, based on the partial order $\prec_{ih}$
(5a) Identify critical points $\sigma^{(k)}$ as local min of critical monotonic sections
(5b) Compute (Un-)stable manifolds and set connections $\sigma^{(k)} \to \sigma^{(k-1)}$
(6) Update the global thickening for the next iteration

6.2 Implementation tricks

Monotonic sections and their extrema In the continuous case, the stable manifold of a critical point $\sigma^{(k)}$ can be decomposed into regions of homogeneous flow, each region corresponding to the intersection of provided that all points belong to the intersection one expects a monotonic regions, that is a non-empty region that represents $W^s(\sigma^{(k)}) \cup W^u(\sigma^{(k+1)})$.

6.3 Details for step 1

Definitions. We assume that:

- The critical points $\sigma^{(k-1)}$ of index $k-1$ have been identified.
- The global thickening $\mathsf{T}[k-1]$ has been computed, with n_k the cardinal of its vertex set.
- The samples in $\mathsf{T}[k-1]$ have been sorted by increasing function value, so that p_1 and p_{n_k} respectively refer to the sample with minimum and maximum function values.

Due to the recursive nature of the sets constructed, we first define:

Definition. 1. *The* bifurcating samples $B_i^{(k)}$ *and the* confluence samples $C_i^{(k)} \subset B_i^{(k)}$ *for* $i = 1, \ldots, n_k$*, are subsets of vertices of the vertex set of the global thickening* $\mathsf{T}[k-1]$*, built incrementally by processing the* n_k *vertices of* $\mathsf{T}[k-1]$*, as indicated below. A sample of* $\mathsf{T}[k-1]$ *not in* $B_{n_k}^{(k)}$ *is termed* regular.

The subgraph of $\mathsf{T}[k-1]$ *induced by the bifurcating samples is called the* global stream $\mathsf{B}[k]$.

▷Initialization. One has:

- $B_i^{(k)}$ This set is initialized with the lowest sample p_i such that two points from its lower link flow to different index $k-1$ critical points in the global thickening $\mathsf{T}[k-1]$. Formally:

$$\omega^-_{|\mathsf{T}[k-1]}(p_j) \neq \omega^-_{|\mathsf{T}[k-1]}(p_k), \text{ with } p_j, p_k \in L^-_{|\mathsf{T}[k-1]}(p_i). \tag{2}$$

- Upon initialization of $B_i^{(k)}$, one set $C_i^{(k)} = \emptyset$.

▷Iteration. Requires three steps:

- Update of the confluence set: $C_i^{(k)} = C_{i-1}^{(k)} \cup \{p_i\}$ with p_i a *confluence sample* characterized by: its entire lower link is a subset of $B_{i-1}^{(k)}$.
- Definition of the *bifurcation lower link* $L^{-\rightsquigarrow}_{|\mathsf{T}[k-1]}(p_i)$ of p_i, by removing from the lower link all the bifurcating samples that are not confluence samples:

$$L^{-\rightsquigarrow}_{|\mathsf{T}[k-1]}(p_i) = \begin{cases} L^-_{|\mathsf{T}[k-1]}(p_i) \backslash (B_{i-1}^{(k)} \backslash C_{i-1}^{(k)}), & \text{if } p_i \notin C_{i-1}^{(k)} \\ L^{-\rightsquigarrow}_{|\mathsf{T}[k-1]}(p_j), \text{ with } p_j \text{ the lowest neighbour of } p_i, & \text{otherwise.} \end{cases} \tag{3}$$

- Classification of p_i as a *bifurcating sample* if:

$$\omega^-_{|\mathsf{T}[k-1]}(p_j) \neq \omega^-_{|\mathsf{T}[k-1]}(p_k) \text{ for at least two neighbors } p_j, p_k \in L^{-\rightsquigarrow}_{|\mathsf{T}[k-1]}(p_i), \tag{4}$$

and update $B_i^{(k)} = B_{i-1}^{(k)} \cup \{p_i\}$ accordingly.

Properties. The following properties are direct consequences of the iterative constructions:

- Confluence samples are only observed for $k \geq 2$.
- The vertex set of $\mathsf{T}[k-1]$ is partitioned into bifurcating samples $B_{n_k}^{(k)}$ and regular samples $C(B_{n_k}^{(k)})$.
- One has $C_{n_k}^{(k)} \subset B_{n_k}^{(k)}$.
- $C_{n_1}^{(1)} = \emptyset$

6.4 Details for step 2

To partition the stream $\mathsf{B}[k]$ into equivalence classes of samples facing identical bifurcations, we define:

Definition. 2. [Hook and Anchor] *A* hook *of index k, also called a k-*hook*, denoted $A^{(k)} = \{\sigma_j^{(k-1)}\}$, is a set of critical points of index $k-1$. The* anchor $A^{(k)}$ *of a sample p_i is the maximal hook satisfying:*

$$\forall \sigma_j^{(k-1)} \in A^{(k)},\ \exists p_j \in L^{-\rightsquigarrow}_{|\mathsf{T}[k-1]}(p_i) \cup \{p_i\} \text{ such that } \omega^{-}_{|\mathsf{T}[k-1]}(p_j) = \sigma_j^{(k-1)}. \tag{5}$$

Anchors are used to group samples as follows:

Definition. 3. [Local thickening and its lower boundary] *The* local thickening $\mathsf{T}(A^{(k)})$ *of a k-anchor $A^{(k)}$ is the induced graph $\mathsf{G}_{|V}$, with V the union of vertices whose anchor is $A^{(k)}$. The* lower boundary *of a local thickening $\partial_l \mathsf{T}(A^{(k)})$ is the set of vertices p_i, for which there exists an edge $(p_i, p_j) \in \mathsf{T}[k-1]$, with $p_j \in L^{-}_{|\mathsf{T}[k-1]}(p_i)$.*

Consider regions of homogeneous flow $W(p, q_i)$, in the neighborhood of the critical point p. Due to the discrete nature of the sampling, we may observe edges in $\mathsf{T}[k-1]$ connecting regions associated with q_i and q_j such that $W(q_i, p) \cap W(q_j, p) = \{p\}$. Such a connexion yields a local thickening, which is spurious since the region $W(q_i, p) \cap W(q_j, p)$ being reduced to a singleton, it cannot define the stable manifold of any critical point. We identify these spurious local thickenings thanks to their *thickness*, and define:

Definition. 4. [Open local thickening] *An interior node of a local thickening is a sample p_i for which all neighboring nodes that are also in $B[k]$ share the same anchor as p_i . A local thickening is* open *if has no interior node.*

Note that vertices in open local thickenings are reassigned as described below. Consider the set $V[\mathsf{T}(\mathsf{A}^{(\mathsf{k})})]$, the reassignment process flows as follows:

- iterate over samples in $V[\mathsf{T}(\mathsf{A}^{(\mathsf{k})})]$ in increasing function value.
- for each sample p consider its neighborhood $q_i \in B_{n_k}^{(k)} \backslash \mathsf{T}(A^{(k)})$

Properties. The following properties holds:

- The vertices of local thickenings partition the vertex set of the steam $\mathsf{B}[k]$.

6.5 Details for step 3

The monotonic sections of a local thickening decompose it into sets of samples having the same α and ω limits for the flow in the local thickening. More formally:

Definition. 5. [Monotonic section] *Given a local minimum α and a local maximum ω of a local thickening $\mathsf{T}(A^{(k)})$, a* monotonic section $\mathsf{M}^{(\alpha,\omega)}(\mathsf{T}(\mathsf{A}^{(\mathsf{k})}))$ *is the sub-graph of $\mathsf{T}(A^{(k)})$ induced by the set of vertices such that $\omega^{-}_{|\mathsf{T}(\mathsf{A}^{(\mathsf{k})})}(p_i) = \alpha$ and $\omega^{+}_{|\mathsf{T}(\mathsf{A}^{(\mathsf{k})})}(p_i) = \omega$.*

In the sequel, for the sake of conciseness, we shall simply write $\mathsf{M_i}(A^{(k)})$ for $\mathsf{M}_{i,j}^{(\alpha_i,\omega_j)}(\mathsf{T}(A^{(k)}))$. To handle the case of monotonic sections consisting of confluence samples, and whose local minima are not candidate for critical points, we define:

Definition. 6. [Linker] *A monotonic section $\mathsf{M_j}(A^{(k)})$ is called a* linker *if:*

$$\forall p_i \in \mathsf{M_j}(A^{(k)}) \setminus \partial_l \mathsf{T}(A^{(k)}), \text{ one has } p_i \in C_n^{(k)}. \tag{6}$$

We finally arrive at the graph encoding incidences between all the monotonic sections of local thickenings at level k (Fig. 1(C)):

Definition. 7. [Incidence graph of monotonic sections] *Two monotonic sections are said to be* incident *if in the stream $\mathsf{B}[k]$ the union of their vertices is connected. They are* interior incident *if they share the same anchor and* exterior incident *otherwise. The incidence graph of monotonic sections $G_M^{(k)}$ is the graph defined as follows: its contains one node per monotonic section which is not a linker; two incident monotonic sections are connected by an edge.*

Properties. The following properties are immediate:

- The vertex set of the set of monotonic sections associated is a covering of the local thickening.
- The vertex set of $G_M^{(k)}$ is a subset of the set of monotonic sections found at $k-st$ iteration, provided that linker monotonic sections are not included.
- The edge set of the union of monotonic sections is included in the edge set of the local thickening.

6.6 Details for step 4

Our goal is to edit $G_M^{(k)}$ into simplified graph $\overline{G}_M^{(k)}$, retaining selected edges only. The process relies on three partial orders $\prec_i$, $\prec_h$, $\prec_{ih}$, the latter mimicking the edges found in the MSW chain complex of a Morse function.

In sections 6.6.1 and 6.6.2, we respectively define the partial orders and perform a cases study of all configurations.

Remark 1. *Because the partial order $\prec_i$ refers to the inclusion between anchors, in the sequel, we are only concerned with the comparisons of monotonic sections associated with different anchors.*

6.6.1 The partial orders

The *inclusion partial order* between two k-anchors yields a partial between monotonic sections, denoted $\prec_i$:

Definition. 8. [Partial order $\prec_i$]

$$\mathsf{M_i}(A_u^{(k)}) \prec_i \mathsf{M_j}(A_v^{(k)}) \iff A_u^{(k)} \subset A_v^{(k)} \tag{7}$$

The *height partial order*, denoted $\prec_h$, aims at comparing monotonic sections linked by a (sequence of) monotonic section(s) within the graph $G_M^{(k)}$:

Definition. 9. [Partial order $\prec_h$] *The partial order is defined by:*

$$\mathsf{M_i}(A_u^{(k)}) \prec_h \mathsf{M_j}(A_v^{(k)}) \iff \exists \mathsf{C} \text{ is a monotonic section with extrema } (\alpha_i, \omega_j) \tag{8}$$

$$\text{where} \begin{cases} \mathsf{C} \subset \mathsf{T}[k-1] \text{ is the graph induced by } \bigcup_{t=0}^{n} V[\mathsf{M_k}(A_t^{(k)})] \\ \text{with } (\mathsf{M_k}(A_t^{(k)}))_{t<n} \in G_M^{(k)} \text{ a path between } \mathsf{M_i}(A_u^{(k)}) \text{ and } \mathsf{M_j}(A_v^{(k)}) \end{cases} \tag{9}$$

Finally the partial order $\prec_{ih}$ is defined by:

Definition. 10. [Partial order $\prec_{ih}$]

$$\mathsf{M_i}(A_u^{(k)}) \prec_{ih} \mathsf{M_j}(A_v^{(k)}) \iff \begin{cases} \mathsf{M_i}(A_u^{(k)}) \prec_i \mathsf{M_j}(A_v^{(k)}) \\ \text{and } \mathsf{M_i}(A_u^{(k)}) \prec_h \mathsf{M_j}(A_v^{(k)}) \end{cases} \tag{10}$$

6.6.2 Classification of edges of $G_M^{(k)}$

Given a partial order $\prec$ and two elements u and v, two situations are faced: either the two elements can be compared ($u \prec v$ or $v \prec u$), or they cannot ($u \parallel v$). In dealing with two partial orders denoted $\prec_i$ and $\prec_h$, this yields five situations, namely: two if the two partial orders apply, two if only one of them applies, and one if none of them applies.

In the following, let us identify edges, say $(\mathsf{M_i}(A_u^{(k)}), \mathsf{M_j}(A_v^{(k)})$ wrt to these two relations:

- *canonical* : $\mathsf{M_j}(A_u^{(k)}) \prec_i \mathsf{M_i}(A_v^{(k)})$, $\mathsf{M_i}(A_u^{(k)}) \prec_h \mathsf{M_j}(A_v^{(k)})$.
- *composite* : $\mathsf{M_j}(A_u^{(k)}) \prec_i \mathsf{M_i}(A_v^{(k)})$, $\mathsf{M_i}(A_v^{(k)}) \prec_h \mathsf{M_j}(A_u^{(k)})$.
- *H-non-I* : $\mathsf{M_j}(A_u^{(k)}) \parallel_i \mathsf{M_i}(A_v^{(k)})$, $\mathsf{M_i}(A_u^{(k)}) \prec_h \mathsf{M_j}(A_v^{(k)})$.
- *I-non-H* : $\mathsf{M_j}(A_u^{(k)}) \parallel_h \mathsf{M_i}(A_v^{(k)})$, $\mathsf{M_i}(A_u^{(k)}) \prec_i \mathsf{M_j}(A_v^{(k)})$.
- *non-hierarchical* : $\mathsf{M_j}(A_u^{(k)}) \parallel_i \mathsf{M_i}(A_v^{(k)})$, $\mathsf{M_i}(A_u^{(k)}) \parallel_h \mathsf{M_j}(A_v^{(k)})$.

In the following, we discuss the situations where these types of edges occur.

Composite edge. This type of incidence id illustrated in Fig. 6. We term *stick* and *fork*, the vertices associated to the composite edge $A_u^{(k)}, A_v^{(k)}$ Notice that the stick might be minimal wrt to $\prec_i$, but in terms of flow, the omega of the *stick* flows into the alpha of the *fork* in the graph induced by the vertex sets of the two monotonic sections. So the union of vertex sets of the stick and the fork form a monotonic region whose extrema are (α_u, ω_v). Note that as far as the *fork* is concerned, as it is not minimal wrt to $\prec_i$ it has at least two incident edges that are canonical, say $\mathsf{M_j}(A_{c_1}^{(k)})$ and $\mathsf{M_j}(A_{c_2}^{(k)})$. In an ideal setting the 'premature' merge of the two c.c. associated to $\mathsf{M_j}(A_{c_1}^{(k)})$ and $\mathsf{M_j}(A_{c_2}^{(k)})$ shouldn't occur and one should redistribute the points in the support of the fork and stick to the incident monotonic sections found downstream.

H-non-I edge. Such an edge occurs in the context of composite edges detailed above, as the stick can also be incident to the $\mathsf{M}_\mathsf{j}(A_{c_1}^{(k)})$ and $\mathsf{M}_\mathsf{j}(A_{c_2}^{(k)})$ and the edge in will be labeled H-non-I.

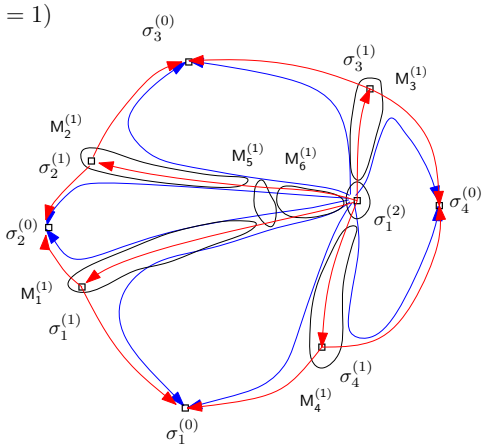
I-non-H edge. This edge occurs when a cycle (Fig. 9) is found in the sub-level associated with a critical value. One way to better represent this type of incidence , which aims at 'correcting' the flow induced at k+1 is the following : consider two adjacent nodes satisfying $\mathsf{M}_\mathsf{i}(A_u^{(k)}) \prec_i \mathsf{M}_\mathsf{i}(A_v^{(k)})$ and $\mathsf{M}_\mathsf{i}(A_u^{(k)}) \parallel_h \mathsf{M}_\mathsf{i}(A_v^{(k)})$ The operation we wish to perform is to shift $\mathsf{M}_\mathsf{i}(A_v^{(k)})$ to some region where all critical points of index k are reachable. In order to do so, we relabel the samples found at the boundary of the two monotonic sections in question.

Non-hierarchical edge. A non-hierarchical edge occurs due to *spurious* connections in the $\mathsf{T}[k-1]$, which may cause erroneous paths when considering the flow operator at the next iteration (Fig. 8).

Note that unlike the forks and the sticks, in this case only the edge in $G_M^{(k)}$ is removed, since discretization is responsible only for the incidence relationship, not for the presence of a spurious monotonic section. However, in the $k+1$th iteration, the edges in $\mathsf{T}[k-1]$ accounting for a non-hierarchical edge must be dismissed.

Figure 6 Incidence between monotonic sections: composite edge. The split of flow lines after yields forks and sticks. **(A)** At $k = 1$, in the discrete setting, pairs and triples of local minima define the monotonic sections $\mathsf{M}_1^{(1)}$ to $\mathsf{M}_6^{(1)}$. In particular, the function value is less $\mathsf{M}_5^{(1)}$ than on $\mathsf{M}_6^{(1)}$ (using the partial order $\prec_h$: $\mathsf{M}_5^{(1)} \prec_h \mathsf{M}_6^{(1)}$; but $\mathsf{M}_6^{(1)}$ bifurcates with respect to two minima while $\mathsf{M}_5^{(1)}$ bifurcates with respect to three minima (using the partial order $\prec_i$: $\mathsf{M}_6^{(1)} \prec_i \mathsf{M}_5^{(1)}$). The two partial orders conflict, so that the edge connecting $\mathsf{M}_5^{(1)}$ and $\mathsf{M}_6^{(1)}$ is called *composite*. Pictorially, the two monotonic sections $\mathsf{M}_5^{(1)}$ and $\mathsf{M}_6^{(1)}$ are respectively called a *fork* and a *stick*. Note that these sections yield confluence samples at $k = 1$. **(B, Inset for A)** Notice that in the neighborhood of $\sigma_1^{(2)}$, no sample from the region $W(\sigma_1^{(2)}, \sigma_2^{(0)})$ flows to $\sigma_2^{(0)}$. Ideally, extensions of $\mathsf{M}_1^{(1)}$ and $\mathsf{M}_2^{(1)}$ should split $\mathsf{M}_5^{(1)}$ and $\mathsf{M}_6^{(1)}$.

(A) $(k = 1)$



(B) $(k = 1)$

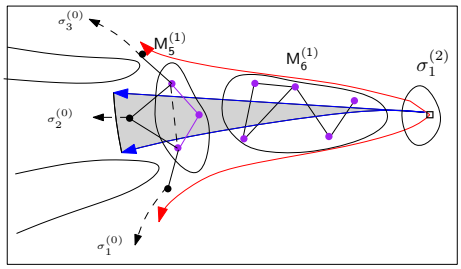


Figure 7 Composite edges and linkers. (A) In the smooth setting, four flow lines (in red) originating at a $\sigma^{(1)}$ converge to the index two critical point (red circle). Each black contoured region $\mathsf{M}_{\mathsf{i}}^{(1)}$ corresponds to a monotonic region at $k = 1$ (samples omitted for clarity). **(B)** In the discrete case, by considering the flow to index one saddles, the analogs of smooth flow lines merge in the support of $\mathsf{M}_5^{(1)}$ and $\mathsf{M}_6^{(1)}$. Note that the purple and turquoise samples are respectively bifurcating and confluence samples. Altogether, these samples, they define the linker $\mathsf{M}_1^{(2)}$.

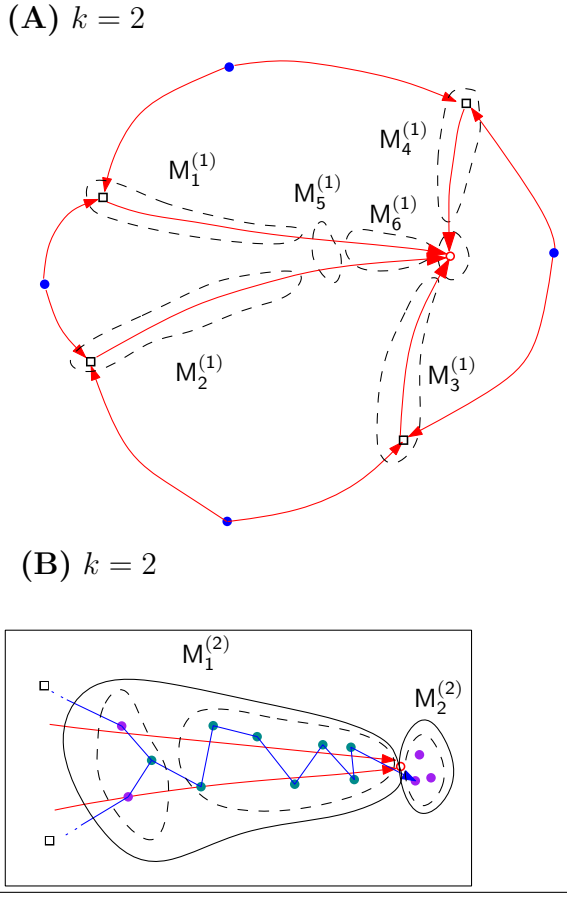


Figure 8 Incidence between monotonic sections: non-Hierarchical edge (A) A selection of monotonic sections(black contoured regions) is displayed together with selected flow lines originating at minima(blue dots) and ending at an index 2 critical points(blue curves). **(B)** At $k=1$, $\mathsf{M}_1^{(1)}$ and $\mathsf{M}_2^{(1)}$ are monotonic sections , whose incidence is non-hierarchical. The red edges (between bifurcating samples– purple) that define this incidence have to be removed when considering the global thickening of index k. The green edges define a canonical incidence between $\mathsf{M}_4^{(1)}$ and $\mathsf{M}_1^{(1)}$, and between $\mathsf{M}_4^{(1)}$ and $\mathsf{M}_3^{(1)}$. **(C)** At $k=2$, the red curves are flow lines originating at index 1 critical points and ending at index 2 critical points. The regular samples (black) who lie in the support of $\mathsf{M}_1^{(1)}$ and $\mathsf{M}_2^{(1)}$ (joined by a dotted red edge) would have formed a linker if we wouldn't have removed the connections between them in the global thickening.

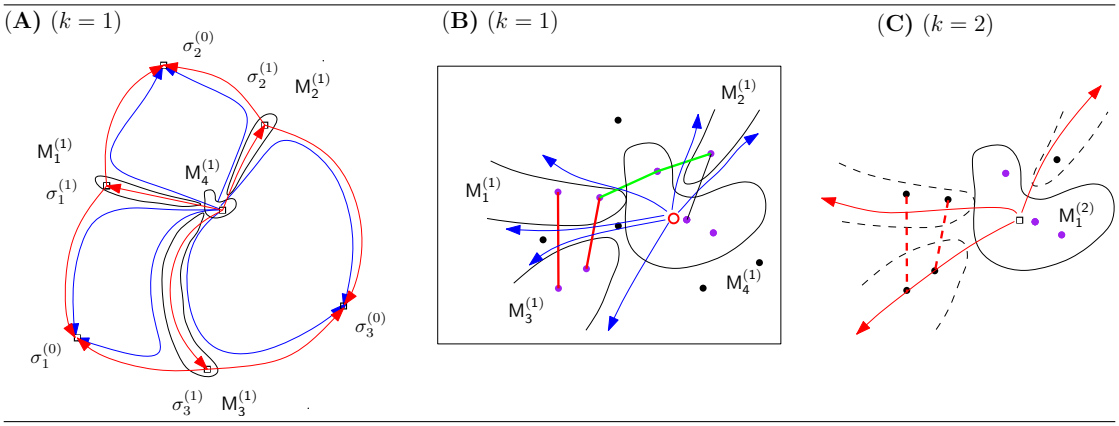
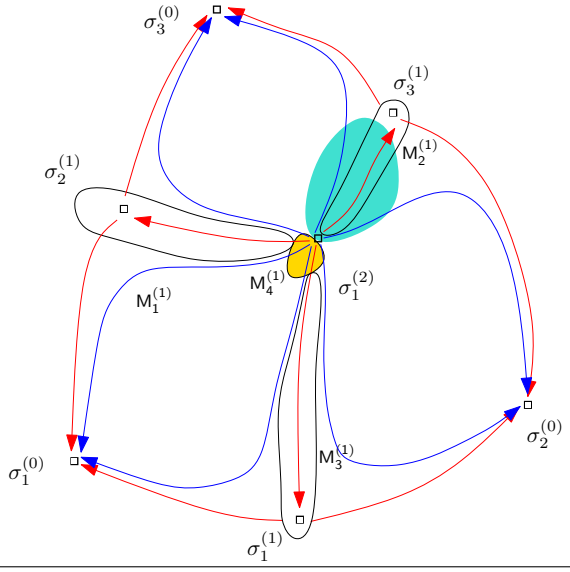


Figure 9 Incidence between monotonic sections: I-non-H edge About an index two critical point $\sigma_1^{(2)}$, a cycle (order one homology) is created upon passing the index one critical point $\sigma_1^{(2)}$. A I-non-H edge is observed between the monotonic section $\mathsf{M}_4^{(1)}$ and $\mathsf{M}_2^{(1)}$ since $\mathsf{M}_2^{(1)} \prec_i \mathsf{M}_4^{(1)}$ but $\mathsf{M}_2^{(1)} \|_h \mathsf{M}_4^{(1)}$. The two monotonic sections cannot be compared for the $\prec_h$ partial order since their union does not form a monotonic section—it has two local minima.



Properties. The following properties hold:

- The vertex set associated to the union of such *fork-stick* couple will be a linker at $k+1$ iteration (Fig. 7).
- The samples that have incident edges in the $\mathsf{T}[k-1]$ that link two monotonic sections sharing a non-hierarchical edge in $G_M^{(k)}$ would determine the presence of an open local thickening or a linker monotonic section at $k+1$ iteration, as illustrated in 8(Left). This is what justifies the removal of edges that link the two monotonic sections in question.
- The vertex set and the edge set of $\overline{G}_M^{(k)}$ are subsets of the vertex set, respectively the edge set of $G_M^{(k)}$.

6.7 Details for step 5

Definition. 11. [Critical point] *A critical point is the local minimum of a minimal monotonic section of the graph* $\overline{G}_M^{(k)}$.

Remark 2. [Linking $\sigma^{(k)}$ to $\sigma^{(k-1)}$] *Each* $\sigma^{(k)}$ *will be linked to the* $\sigma^{(k-1)}$ *that form the anchor of its associated monotonic section.*

Definition. 12. [Stable manifold of $\sigma^{(k-1)}$] *The stable manifold of* $\sigma^{(k-1)}$ *is defined by:*

$$W^s(\sigma^{(k-1)}) = \{p \in C(B[k]) \cup \bigcup \mathsf{M}_\mathsf{l}(A^{(k)}) | \omega^-_{|\mathsf{T}[\mathsf{k}-1]}(p) = \sigma^{(k-1)}\}, \tag{11}$$

where $\mathsf{M}_\mathsf{l}(A^{(k)})$ *are linkers and* $C(B[k])$ *is the complementary of* $V[\mathsf{T}[\mathsf{k}-1]]$.

Remark 3. *When linkers (and implicitly confluence samples) are present, the stable manifold of a critical point is not only described by the points belonging to* $C(B_k^{(n_k)})$.

6.8 Details for step 6

The global thickening $\mathsf{T}[k]$ to be used during the next iteration is defined as follows:

Definition. 13. [Global thickening] *The vertex set of* $\mathsf{T}[k]$ *is defined by:*

$$V[\mathsf{T[k]}] = \bigcup_j V[\mathsf{M_i}(\mathsf{A_j^{(k)}})] \text{ , with } \mathsf{M_i}(A_j^{(k)}) \in G_M^{(k)} \tag{12}$$

The edge set of $\mathsf{T}[k]$ *is defined by:*

$$E[\mathsf{T[k]}] = \bigcup_j E[\mathsf{M_i}(\mathsf{A_j^{(k)}})] \cup C, \text{ with} \tag{13}$$

$$C = \{e = (u,v) \in \mathsf{T}[k-1] | u \in \mathsf{M_i}(A_u^{(k)}), v \in \mathsf{M_j}(A_v^{(k)}) \text{ and } (\mathsf{M_i}(A_u^{(k)}) \prec_h \mathsf{M_j}(A_v^{(k)}))\} \tag{14}$$

7 Benchmark

⟶Frederic says: I suggest to organize the dataset as follows, since we may not include the Hasse diagram in the certif paper←

```
\subsection{2D terrains}
\subsubsection Equations and Illustrations
\subsubsection Hasse diagrams
```

7.1 Equations

Table 1 2D function used in this study

▷ Himmelblau, see also `http://en.wikipedia.org/wiki/Himmelblau%27s_function`.

$$f(x,y) = (x^2 + y - 11)^2 + (x + y^2 - 7)^2 \tag{15}$$

▷ Gauss6a

$$\begin{aligned} f(x,y) = & \frac{1}{20\pi} exp - \frac{(x-25)^2 + \frac{(y-17)^2}{3}}{2} - \frac{1}{24\pi} exp - \frac{\frac{(x-25)^2}{8} + \frac{(y-10)^2}{10}}{2} - \frac{1}{16\pi} exp - \frac{\frac{(x-25)^2}{10} + \frac{(y-20)^2}{8}}{2} \\ & - \frac{1}{6\pi} exp - \frac{(x-20)^2 + \frac{(y-25)^2}{3}}{2} + \frac{1}{12\pi} exp - \frac{\frac{(x-30)^2}{3} + \frac{(y-19)^2}{2}}{2} + \frac{1}{20\pi} exp - \frac{\frac{(x-20)^2}{5} + \frac{(y-22)^2}{2}}{2} \end{aligned} \tag{16}$$

▷ Gauss6b

$$\begin{aligned} f(x,y) = & \frac{1}{20\pi} exp - \frac{(x-25)^2 + \frac{(y-17)^2}{3}}{2} - \frac{1}{24\pi} exp - \frac{\frac{(x-25)^2}{8} + \frac{(y-10)^2}{10}}{2} - \frac{1}{16\pi} exp - \frac{\frac{(x-25)^2}{10} + \frac{(y-20)^2}{8}}{2} \\ & - \frac{1}{6\pi} exp - \frac{(x-20)^2 + \frac{(y-25)^2}{3}}{2} + \frac{1}{12\pi} exp - \frac{\frac{(x-30)^2}{3} + \frac{(y-19)^2}{2}}{2} + \frac{1}{20\pi} exp - \frac{\frac{(x-20)^2}{2} + \frac{(y-17)^2}{5}}{2} \end{aligned} \tag{17}$$

▷ Rastrigin, see also also `http://en.wikipedia.org/wiki/Rastrigin_function`.

$$f(\mathbf{x}) = An + \sum_{i=1}^{d} \left(x_i^2 + A\cos(2\pi x_i)\right), (\mathbf{x} = (x_1, \ldots, x_d)) \tag{18}$$

Table 2 3D functions used in this study

▷ poly1

$$f(x,y,z) = \left(x^2+y^2+z-1\right)^2 + \left(x^2+z^2+y-2\right)^2 + \left(y^2+z^2+x-3\right)^2 \tag{19}$$

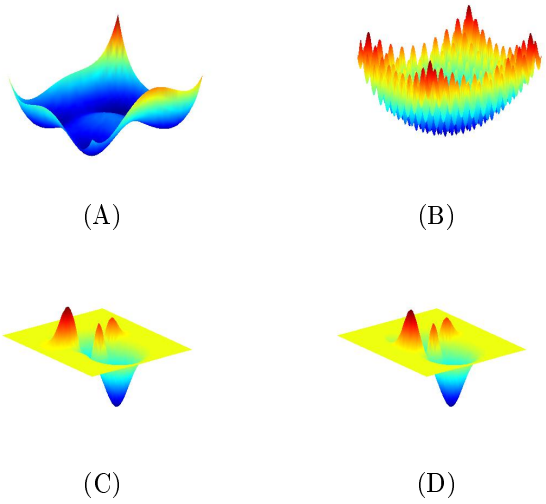
▷ poly2

$$f(x,y,z) = \left(x^2+y^2+z-5\right)^2 + \left(x^2+z^2+y-7\right)^2 + \left(y^2+z^2+x-11\right)^2 \tag{20}$$

▷ poly3

$$f(x,y,z) = \left(x^2+y^2+z-5\right)^2 + \left(x^2+z^2+y-2\right)^2 + 7 + x^2 + 0.2\,x \tag{21}$$

Figure 10 Bi-variate functions of the benchmark: illustrations (**A**) Himmelblau (picture from wikipedia) (**B**) 2D-Rastrigin (picture from wikipedia) (**C**) Gaussian-6A (**D**) Gaussian-6B



7.2 3D Functions

Polynomial P1 We consider the polynomial of Eq. (19). Using the tools presented in section 3.1, it can be shown that this polynomial has 11 critical points. Ordering them by increasing critical value yields the following sequence of indices: four index 0, four index 1 saddles, one index 2 saddle, one index 1 saddle, one index 2 saddle.

The critical points and associated Hasse diagram is presented on Fig. 11, while the level set surfaces and the accompanying Betti numbers of the sub-level set $M_{c+\varepsilon}$ are depicted on Fig. 12 .

Figure 11 poly1(Eq. (19)): overview of critical points and Morse-Smale diagram

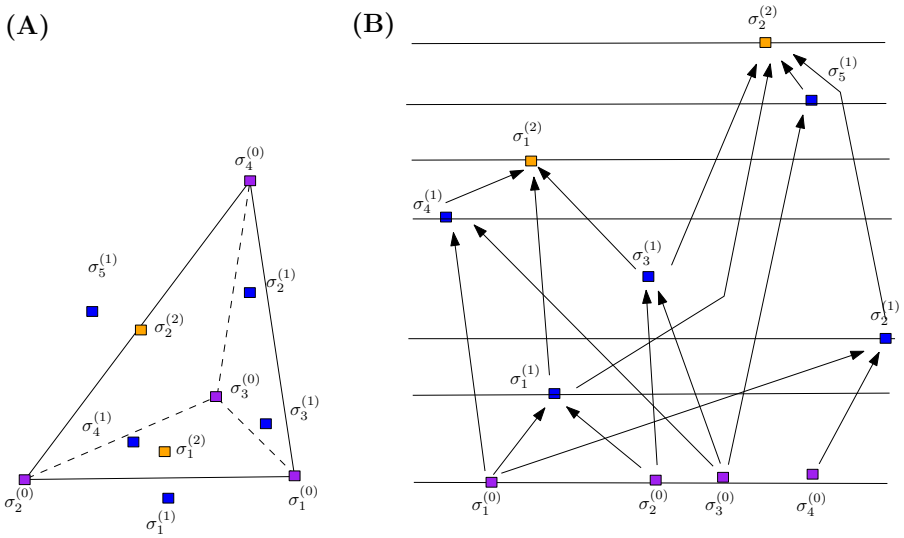
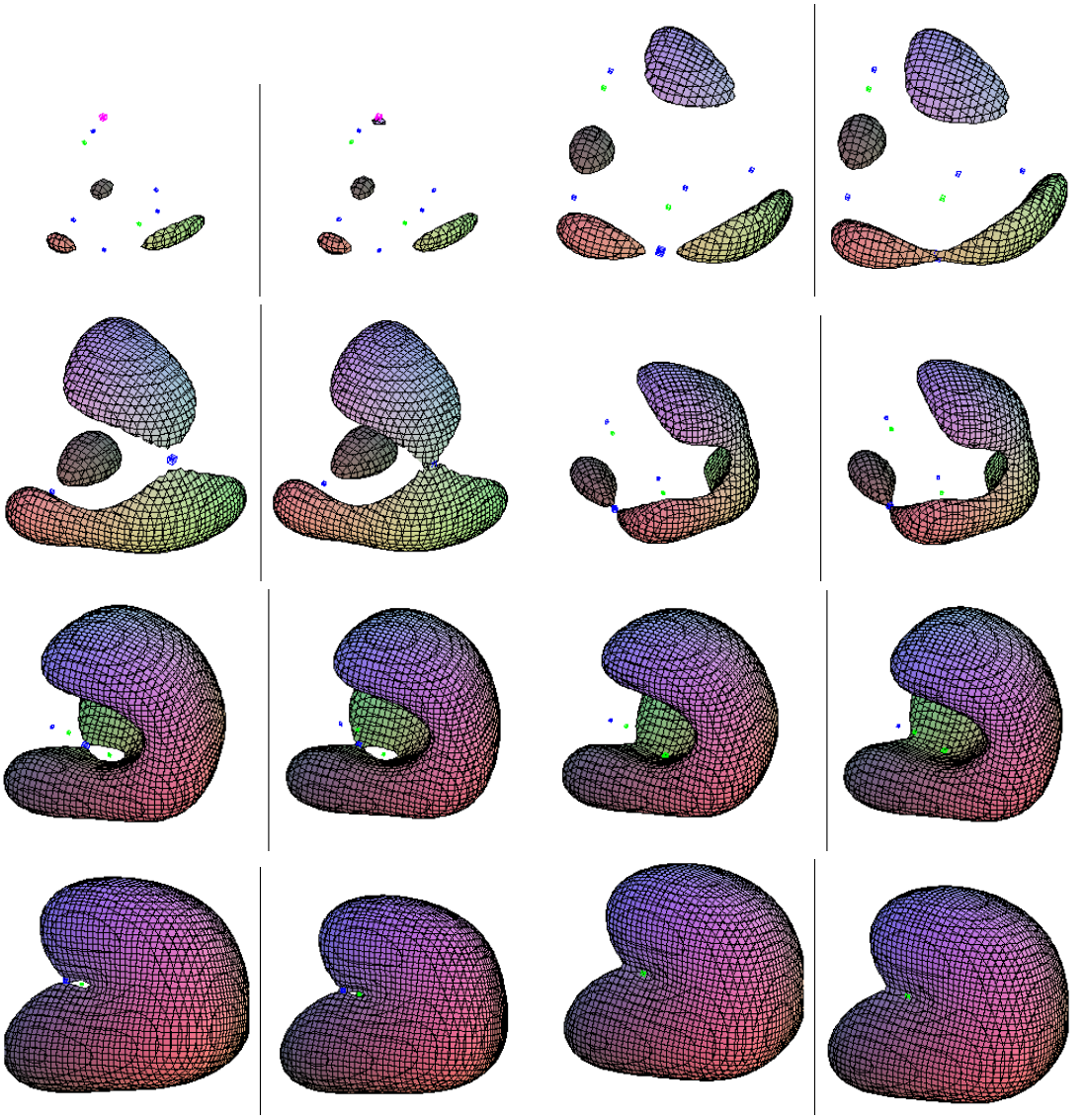
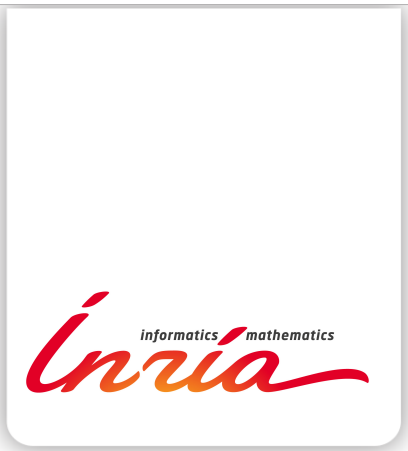


Figure 12 Plots of the sub-level sets of the polynomial poly1(Eq. (19) before and after passing a critical point. The eight pairs of plots correspond to passing the critical points with id 4 to 11 (critical points #4, #6, #8, #10 for column (a), and #5,#7, #9, #11 for column (b). The evolution of the Betti numbers for the 11 sub-level sets $M_{c+\varepsilon}$ are: #1, index 0 : (1,0,0) #2 index 0 : (2,0,0) #3, index 0 : (3,0,0) #4, index 0 : (4,0,0) #5, index 1 : (3,0,0) #6, index 1 : (2,0,0) #7, index 1 : (1,0,0) #8, index 1 : (1,1,0) (creates a filled torus) #9, index 2 : (1,0,0) (fills the void in the middle of the filled torus) #10, index 1 : (1,1,0) (recreates a filled torus) #11 index 2 : (1,0,0) (refills the void of the filled torus)



(a) (b)



RESEARCH CENTRE
SOPHIA ANTIPOLIS – MÉDITERRANÉE

2004 route des Lucioles - BP 93
06902 Sophia Antipolis Cedex

Publisher
Inria
Domaine de Voluceau - Rocquencourt
BP 105 - 78153 Le Chesnay Cedex
inria.fr

ISSN 0249-6399